Chapter 2

Vector Spaces - An Introduction

A *vector space* over a scalar field $\mathcal{F}$ (in our work, we use both the real numbers $\mathcal{R}$ or the complex numbers $\mathcal{C}$ as scalars) is a nonempty set of elements, called *vectors*, with two laws of combination: vector addition and scalar multiplication. Vector addition must satisfy

1. To every $\vec{X}$ and $\vec{Y}$ in vector space **V**, there is a unique sum vector $\vec{Z} \in$ **V** such that $\vec{Z} = \vec{X} + \vec{Y}$ (*i.e.*, the space is closed under vector addition).
2. Vector addition is associative: $(\vec{X} + \vec{Y}) + \vec{Z} = \vec{X} + (\vec{Y} + \vec{Z})$.
3. A zero vector exists with the property $\vec{0} + \vec{X} = \vec{X}$ for all $\vec{X} \in$ **V**. Vector $\vec{0}$ is the identity element under vector addition.
4. For each $\vec{X} \in$ **V** there exists $-\vec{X} \in$ **V** such that $\vec{X} + (-\vec{X}) = \vec{0}$.
5. Addition is commutative: $\vec{X} + \vec{Y} = \vec{Y} + \vec{X}$.

Scalar multiplication must satisfy

6. To every scalar $\alpha \in \mathcal{F}$ and vector $\vec{X} \in$ **V** there is a unique vector $\alpha\vec{X} \in$ **V** (*i.e.*, the space is closed under scalar multiplication).
7. Scalar multiplication is associative: $\alpha(\beta\vec{X}) = (\alpha\beta)\vec{X}$ for all $\vec{X} \in$ **V** and $\alpha, \beta \in \mathcal{F}$.
8. Scalar multiplication is distributive with respect to vector addition: $\alpha(\vec{X} + \vec{Y}) = \alpha\vec{X} + \alpha\vec{Y}$.
9. Scalar multiplication is distributive with respect to scalar addition: $(\alpha+\beta)\vec{X} = \alpha\vec{X} + \beta\vec{X}$.
10. The unit element $1 \in \mathcal{F}$ is the identity element for scalar multiplication: $1\vec{X} = \vec{X}$ for all $\vec{X} \in$ **V**.

<u>Linear Dependence/Independence</u>

Vectors $\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n$ are said to be *linearly independent* if there exists $\alpha_i \in \mathcal{F}$, not all zero, with the property $\alpha_1\vec{X}_1 + \alpha_2\vec{X}_2, \ldots + \alpha_n\vec{X}_n = \vec{0}$. That is, the vectors are linearly dependent (or dependent) if a non-trivial linear combination sums to the zero vector. If $\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n$ are not linearly dependent, they are said to be *linearly independent* (or independent). If $\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n$ are independent, then $\alpha_1\vec{X}_1 + \alpha_2\vec{X}_2, \ldots + \alpha_n\vec{X}_n = \vec{0}$ implies that $\alpha_1 = \alpha_2 = \ldots = \alpha_n = 0$.

If the maximal number of independent vectors in space **V** is finite (say n), then **V** is said to be *finite dimensional*, and n is said to be the *dimension* of **V**. There are many important vector spaces that are not finite dimensional. The general theory of Fourier series, and much of communication and control theory, has an infinite dimensional space as its natural setting. In this course, we consider only finite dimensional spaces over the real number field or the complex number field.

Span

Let A denote any subset of vector space **V**. The set *spanned* by A consists of all possible linear combinations of vectors in A. It is denoted by Span(A).

Notes:

1) $A \subset \text{Span}(A)$

2) If A and B are sets of vectors with $A \subset B$ then $\text{Span}(A) \subset \text{Span}(B)$.

3) If a finite set $\vec{X}_1, \vec{X}_2, \ldots, \vec{X}_n$ spans **V** then every linearly independent set contains at most n vectors.

A linearly independent set that spans **V** is called a *basis* of **V**.

Notes:

1) If a vector space **V** has a basis with n elements, then all other bases have the same number n of vectors. Integer n is said to be the *dimension* of the vector space.

2) In an n-dimensional space, any set containing n+1 or more vectors is linearly dependent.

3) In a finite dimensional space, any linearly independent set of vectors can be extended to a basis.

Vector Representation

Let $\vec{\alpha}_1, \vec{\alpha}_2, \ldots, \vec{\alpha}_n$ denote a basis of **V**. Let $\vec{X}$ be any vector in **V**. We can represent uniquely

$$\vec{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} x_i \vec{\alpha}_i = \left[\vec{\alpha}_1 \; \vdots \; \vec{\alpha}_2 \; \vdots \; \cdots \; \vdots \; \vec{\alpha}_n \right] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{2-1}$$

for some set of scalars $x_1, x_2, \ldots, x_n$. The scalars $x_1, x_2, \ldots, x_n$ are called the *coordinates* of $\vec{X}$. In what follows, coordinates will be denoted by italicized, lower case letters. The quantity $[x_1, x_2, \ldots, x_n]^T$ is a *coordinate vector*; it can only be used in (2-1) to represent vector $\vec{\mathrm{X}}$ once an underlying basis has been identified.

This is where the notation gets a little confusing. We write $\vec{\mathrm{X}} = [\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n]^T$, where x_k, $1 \le \mathrm{k} \le \mathrm{n}$, are the *components* of $\vec{\mathrm{X}}$. On the other hand, we write $\vec{\mathrm{X}}$ as in (2-1) where $[x_1, x_2, \ldots, x_n]^T$ is a vector of *coordinates* (remember: Italicized, lower-case letters for *coordinates* and unitalicized, lower-case letters for *components*).

<u>Subspace</u>

A *subspace* **W** of **V** is a nonempty subset of **V** which is itself a vector space. The subspace must be over the same scalar field $\mathcal{F}$ as **V**.

<u>Intersection of Subspaces</u>

Let $\mathbf{W_1}$ and $\mathbf{W_2}$ be subspaces of **V**. Then we define the intersection of $\mathbf{W_1}$ and $\mathbf{W_2}$ as

$$\mathbf{W_1} \cap \mathbf{W_2} = \left[\vec{\mathrm{X}} \in \mathbf{V} : \vec{\mathrm{X}} \in \mathbf{W_1} \text{ and } \vec{\mathrm{X}} \in \mathbf{W_2} \right]. \tag{2-2}$$

The intersection of *any collection* of subspaces is a subspace.

Let *A* be any subset of **V**, not necessarily a subspace. There exists subspaces $\mathbf{W_i} \subset \mathbf{V}$ which contain subset *A*; that is, $A \subset \mathbf{W_i} \subset \mathbf{V}$. The intersection of all subspaces which contain *A* is denoted as

$$\bigcap_{A \subset \mathbf{W_i}} \mathbf{W_i}\,. \tag{2-3}$$

It is the smallest (in terms of dimension) subspace that contains *A*. Also, Span(*A*) is the smallest subspace that contains the subset *A*; we write

$$\bigcap_{A \subset \mathbf{W_i}} \mathbf{W_i} = \mathrm{Span}(A)\,. \tag{2-4}$$

Sum of Subspaces

Let $\mathbf{W_1}$ and $\mathbf{W_2}$ denote subspaces of $\mathbf{V}$. Define $\mathbf{W_1} + \mathbf{W_2}$ as the set of all $\vec{X}_1 + \vec{X}_2$, where $\vec{X}_1 \subset \mathbf{W_1}$ and $\vec{X}_2 \subset \mathbf{W_2}$. The sum $\mathbf{W_1} + \mathbf{W_2}$ of subspaces is a subspace of $\mathbf{V}$, and

$$\mathbf{W_1} + \mathbf{W_2} = \text{Span}(\mathbf{W_1} \cup \mathbf{W_2}), \tag{2-5}$$

where

$$\mathbf{W_1} \cup \mathbf{W_2} = \left[\vec{X} \in \mathbf{V} : \vec{X} \in \mathbf{W_1} \text{ or } \vec{X} \in \mathbf{W_2}\right]. \tag{2-6}$$

That is, $\mathbf{W_1} + \mathbf{W_2}$ is the smallest subspace that contains both $\mathbf{W_1}$ and $\mathbf{W_2}$. ***Note that*** $\mathbf{W_1} \cup \mathbf{W_2}$ ***is <u>not</u> a subspace, in general.***

Basis of Subspace

Given an m-dimensional subspace $\mathbf{W}$ in an n-dimensional space $\mathbf{V}$, let

$$\vec{\alpha}_1, \vec{\alpha}_2, \ldots, \vec{\alpha}_m$$

be a basis of subspace $\mathbf{W}$. By adding independent vectors, this basis can be extended to

$$\vec{\alpha}_1, \vec{\alpha}_2, \ldots, \vec{\alpha}_m, \vec{\alpha}_{m+1}, \ldots, \vec{\alpha}_n,$$

a basis of $\mathbf{V}$. That is, a basis for a subspace can be extended (by adding independent vectors) to a basis for the whole space $\mathbf{V}$.

Equality of Subspaces

Let $\mathbf{W_1}$ and $\mathbf{W_2}$ be subspaces of $\mathbf{V}$. We say that $\mathbf{W_1}$ and $\mathbf{W_2}$ are *equal* if $\mathbf{W_1} \subset \mathbf{W_2}$ and $\mathbf{W_2} \subset \mathbf{W_1}$. Alternatively, $\mathbf{W_1} = \mathbf{W_2}$ if both subspaces are spanned by the same finite dimensional basis. Finally, if subspaces $\mathbf{W_1}$ and $\mathbf{W_2}$ have the same finite dimension m, and $\mathbf{W_1} \subset \mathbf{W_2}$, then $\mathbf{W_1} = \mathbf{W_2}$.

Dimension of Sum

Let $\mathbf{W_1}$ and $\mathbf{W_2}$ be subspaces of finite dimensional $\mathbf{V}$. Then

$$\dim\{\mathbf{W_1} + \mathbf{W_2}\} = \dim\{\mathbf{W_1}\} + \dim\{\mathbf{W_2}\} - \dim\{\mathbf{W_1} \cap \mathbf{W_2}\}. \tag{2-7}$$

EXAMPLE

Consider the vector space $\mathcal{R}^3$. Define subspaces $\mathbf{W_1}$ = Span{$[1\ \ 0\ \ 2]^T$, $[1\ \ 2\ \ 2]^T$} and $\mathbf{W_2}$ = Span{$[1\ \ 1\ \ 0]^T$, $[0\ \ 1\ \ 1]^T$}. Find a basis for $\mathbf{W_1} \cap \mathbf{W_2}$. Note that $\dim[\mathbf{W_1}] = \dim[\mathbf{W_2}] = 2$ so $\mathbf{W_1}$ and $\mathbf{W_2}$ are planes in $\mathcal{R}^3$. Now, $\mathbf{W_1}$ and $\mathbf{W_2}$ both contain the origin, so $\mathbf{W_1} \cap \mathbf{W_2}$ is not null. In fact, $\mathbf{W_1} \cap \mathbf{W_2}$ is a line. Since $\mathbf{W_1} \subset \mathbf{W_1} + \mathbf{W_2} \subset \mathcal{R}^3$ we have $2 \le \dim(\mathbf{W_1} + \mathbf{W_2}) \le 3$. By (2-7), we have $4-3 = 1 \le \dim(\mathbf{W_1} \cap \mathbf{W_2}) \le 4-2 = 2$. Now, any $\vec{X} \in \mathbf{W_1} \cap \mathbf{W_2}$ must be expressible in the form

$$\begin{aligned}\vec{X} &= a\begin{bmatrix}1 & 0 & 2\end{bmatrix}^T + b\begin{bmatrix}1 & 2 & 2\end{bmatrix}^T \\ &= c\begin{bmatrix}1 & 1 & 0\end{bmatrix}^T + d\begin{bmatrix}0 & 1 & 1\end{bmatrix}^T\end{aligned}, \tag{2-8}$$

which requires that

$$\begin{aligned}a + b &= c \\ 2b &= c + d \\ 2a + 2b &= d\end{aligned} \tag{2-9}$$

In terms of the variable a, the solution (2-9) is

$b = -3a$

$$c = -2a \tag{2-10}$$
$$d = -4a$$

Finally, we have the representation $\vec{X} = a[1\ \ 0\ \ 2]^T - 3a[1\ \ 2\ \ 2]^T = a[-2\ \ -6\ \ -4]^T$. Hence, the subspace $\mathbf{W_1} \cap \mathbf{W_2} = \text{Span}\{[-2\ \ -6\ \ -4]^T\} = \text{Span}\{[1\ \ 3\ \ 2]^T\}$.

An important special case occurs when $\dim(\mathbf{W_1} \cap \mathbf{W_2}) = 0$. When $\mathbf{W_1} \cap \mathbf{W_2}$ contains only the zero vector we say that the sum $\mathbf{W_1} + \mathbf{W_2}$ is *direct*. The direct sum of the two subspaces is denoted as $\mathbf{W_1} \oplus \mathbf{W_2}$. For each $\vec{X} \in \mathbf{W_1} \oplus \mathbf{W_2}$, $\vec{X}$ has the *unique* decomposition $\vec{X} = \vec{X}_1 + \vec{X}_2$, where $\vec{X}_1 \in \mathbf{W_1}$ and $\vec{X}_2 \in \mathbf{W_2}$.

Direct sums can be extended to a finite number of subspaces. The sum of $\mathbf{W_1}, \mathbf{W_2}, \ldots, \mathbf{W_k}$ is *direct* if for each i, $1 \le i \le k$, we have

$$\mathbf{W_i} \cap \left(\sum_{j \ne i} \mathbf{W_j} \right) = \{\vec{0}\} \tag{2-11}$$

In this case we write $\mathbf{W_1} \oplus \mathbf{W_2} \oplus \mathbf{W_3} \oplus \ldots \oplus \mathbf{W_k}$. For the sum of $\mathbf{W_1}, \mathbf{W_2}, \ldots, \mathbf{W_k}$ to be direct, it is necessary and sufficient for $\dim\{\mathbf{W_1} + \mathbf{W_2} + \ldots + \mathbf{W_k}\} = \dim\{\mathbf{W_1}\} + \dim\{\mathbf{W_2}\} + \ldots + \dim\{\mathbf{W_k}\}$. If $\vec{X} \in \mathbf{W_1} \oplus \mathbf{W_2} \oplus \mathbf{W_3} \oplus \ldots \oplus \mathbf{W_k}$, then we have a *unique* decomposition

$$\vec{\mathbf{X}} = \sum_{i=1}^{k} \vec{\mathbf{X}}_\mathbf{i}\ , \tag{2-12}$$

where $\vec{X}_i \in \mathbf{W_i}$.

If $\mathbf{W_1}$ is a subspace of $\mathbf{V}$, then there exists subspace $\mathbf{W_2}$ such that $\mathbf{V} = \mathbf{W_1} \oplus \mathbf{W_2}$; however, $\mathbf{W_2}$ is not unique.

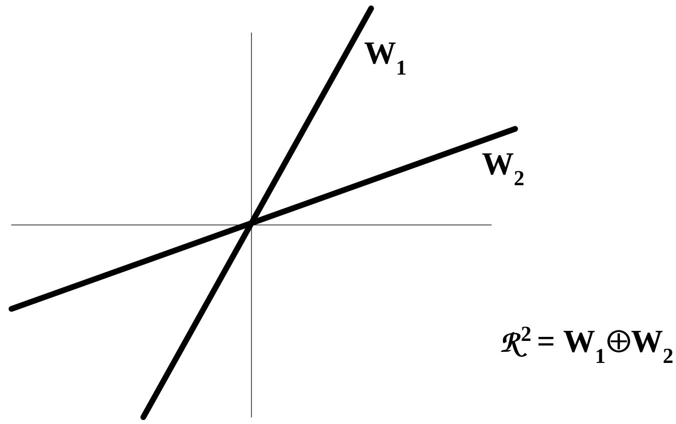

Figure 2-1: Two subspaces whose direct sum is $\mathcal{R}^2$.

Linear Variety

Let **W** be a subspace of vector space **V**, and let $\vec{X}_0 \in$ **V** but $\vec{X}_0 \notin$ **W**. Then the set

$$\{\ \vec{Y} \in \mathbf{V} : \vec{Y} = \vec{X} + \vec{X}_0,\ \vec{X} \in \mathbf{W}\ \} \tag{2-13}$$

is a *translation* of subspace **W**. Such a translation is called a *linear variety* (often called a *linear manifold* and sometimes called a *flat*}; it is **not** a subspace.

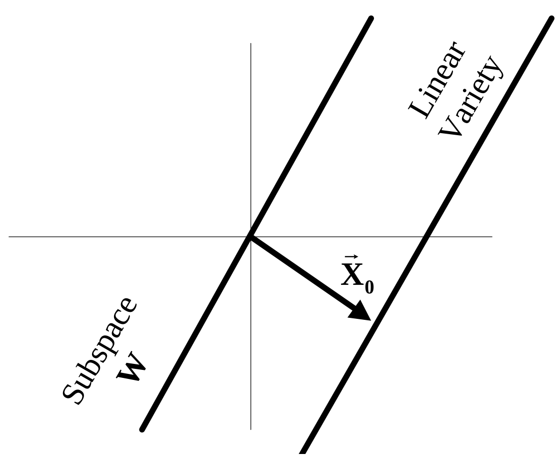

Figure 2-2: Subspace and a linear variety.

Inner Product of Vectors

Let $\vec{X} = [x_1 \;\; x_2 \;\; \ldots \;\; x_n]^T$ and $\vec{Y} = [y_1 \;\; y_2 \;\; \ldots \;\; y_n]^T$ be n-dimensional vectors, which are generally complex valued. The *inner product* (also known as *dot product*) of $\vec{X}$ and $\vec{Y}$ is defined as

$$\left\langle \vec{X}, \vec{Y} \right\rangle = \sum_{k=1}^{n} x_k \bar{y}_k = \vec{Y}^* \vec{X}, \tag{2-14}$$

where $\bar{y}_k$ denotes the complex conjugate of y_k, and $\vec{Y}^*$ denotes the transpose conjugate of $\vec{Y}$. Let $\vec{X}_1$, $\vec{X}_2$, $\vec{Y}_1$, $\vec{Y}_2$ be n-vectors, and let α, β be scalars in $\mathcal{F}$. Then, we have

$$\left\langle \alpha\vec{X}_1 + \beta\vec{X}_2, \vec{Y} \right\rangle = \alpha\left\langle \vec{X}_1, \vec{Y} \right\rangle + \beta\left\langle \vec{X}_2, \vec{Y} \right\rangle, \tag{2-15}$$

that is, the inner product is *linear in its first entry*. Also, we have

$$\left\langle \vec{X}_1, \alpha\vec{Y}_1 + \beta\vec{Y}_2 \right\rangle = \bar{\alpha}\left\langle \vec{X}_1, \vec{Y}_1 \right\rangle + \bar{\beta}\left\langle \vec{X}_1, \vec{Y}_2 \right\rangle, \tag{2-16}$$

that is, the inner product is *conjugate linear in its second entry*.

In the literature, the definition of, and notation for, the inner product varies. For example, MatLab defines the function dot($\vec{X}$,$\vec{Y}$) = $\vec{X}^T\vec{Y} = \vec{Y}^T\vec{X}$. That is, in MatLab, the complex conjugate plays no role in the definition of inner product. In term of MatLab's definition, our definition is

$$\left\langle \vec{X}, \vec{Y} \right\rangle = \text{dot}(\; \vec{X},\; \text{conj}(\vec{Y})\;), \tag{2-17}$$

where conj($\vec{Y}$) is MatLab's conjugate operation.

Orthogonal Complement of Subspace

Let **W** denote a subspace of vector space **V**. Denote the *orthogonal complement* of **W** as $\mathbf{W}^{\perp}$ (called "**W** purp") as

$$\mathbf{W}^{\perp} = \left[\vec{X} \in \mathbf{V} : \left\langle \vec{X}, \vec{Y} \right\rangle = 0 \text{ for all } \vec{Y} \in \mathbf{W} \right]. \tag{2-18}$$

$\mathbf{W}^{\perp}$ is a subspace, and vector space **V** can be decomposed as

$$\mathbf{V} = \mathbf{W} \oplus \mathbf{W}^{\perp}. \tag{2-19}$$

We say that **W** and $\mathbf{W}^{\perp}$ are *orthogonal complements* since they are orthogonal to one another and they decompose the space **V** according to (2-19).

Vector Norms

Generally, a *vector norm* is a function $\| \; \| : \mathbf{V} \rightarrow \mathcal{R}$ that satisfies

1) $\|\vec{X}\| \geq 0$ for all $\vec{X} \in \mathbf{V}$ and $\|\vec{X}\| = 0$ if and only if $\vec{X} = \vec{0}$.
2) $\|\vec{X} + \vec{Y}\| \leq \|\vec{X}\| + \|\vec{Y}\|$ (this is the *triangle inequality*)
3) $\|\alpha\vec{X}\| = |\alpha| \, \|\vec{X}\|$ for all $\vec{X} \in \mathbf{V}$ and α in $\mathcal{F}$.

A vector norm is a way to assign "size" or "length" to a vector.

A useful class of vector norms is the *p-norms* defined by

$$\|\vec{X}\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{1/p}, \quad p \geq 1, \tag{2-20}$$

where $\vec{X} = [x_1 \;\; x_2 \;\; \ldots \;\; x_n]^T$. Of this class of norms, the most useful are

$$\|\vec{X}\|_1 = |x_1| + |x_2| + \cdots + |x_n|$$

$$\|\vec{X}\|_2 = \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2} = \sqrt{\langle \vec{X}, \vec{X} \rangle} \quad . \tag{2-21}$$

$$\|\vec{X}\|_\infty = \max_{1 \le i \le n} |x_i| \quad \text{(called the "infinity norm")}$$

The 2-norm $\|\vec{X}\|_2$ is often called the *Euclidean norm*. **In what follows, the 2-norm is the default. If the kind of norm is not specified in a result, assume that the 2-norm is used.**

The p-norms satisfy the *Holder Inequality*

$$\sum_{i=1}^{n} |x_i \overline{y}_i| \le \|\vec{X}\|_p \|\vec{Y}\|_q, \quad \text{where } \frac{1}{p} + \frac{1}{q} = 1, \tag{2-22}$$

and $\overline{y}_i$ denotes the complex conjugate of y_i. For p = 2, the Holder Inequality is known as the *Cauchy-Schwarz Inequality*.

THEOREM 1 (Cauchy-Schwarz)

Let $\vec{X}$ and $\vec{Y}$ be arbitrary vectors in vector space **V**. Then (remember: use the 2-norm if not stated otherwise)

$$\left| \langle \vec{X}, \vec{Y} \rangle \right| \le \|\vec{X}\| \, \|\vec{Y}\| \tag{2-23}$$

The Cauchy-Schwarz inequality assigns an upper bound on the inner product.

Proof:

The result is obvious if $\vec{X} = \vec{0}$ and/or $\vec{Y} = \vec{0}$. Hence, assume nonzero vectors in this proof. Let α be a real-valued variable and write

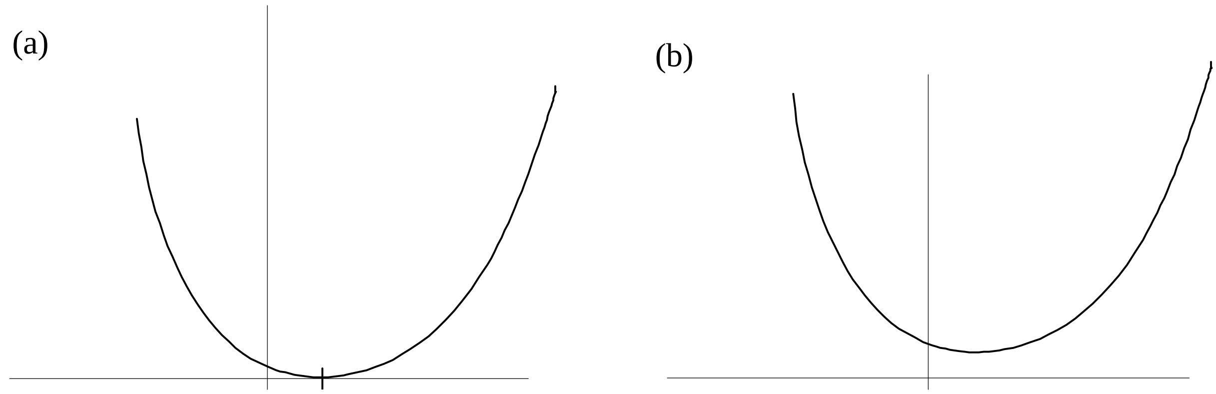

Figure 2-3: Plots of $f(\alpha) = \left\|\vec{X}\right\|^2\alpha^2 + 2\text{Re}[\langle\vec{X},\vec{Y}\rangle]\alpha + \left\|\vec{Y}\right\|^2$**.** Fig. (a) depicts two real equal roots at α_0. Fig. (b) implies that the roots are complex valued.

$$\begin{aligned}\left\|\alpha\vec{X}+\vec{Y}\right\|^2 &= \langle\alpha\vec{X}+\vec{Y},\alpha\vec{X}+\vec{Y}\rangle = (\alpha\vec{X}^*+\vec{Y}^*)(\alpha\vec{X}+\vec{Y}) \\ &= \left\|\vec{X}\right\|^2\alpha^2 + 2\text{Re}[\langle\vec{X},\vec{Y}\rangle]\alpha + \left\|\vec{Y}\right\|^2 \ge 0\end{aligned} \quad . \tag{2-24}$$

This is quadratic in real α; plotted as a function of α, it opens upwards and only touches, or lies entirely above, the α axis. Hence, the quadratic in (2-24) has only real equal roots or complex-valued roots. As a result, the discriminate is nonpositive. That is

$$4\,\text{Re}\{\langle\vec{X},\vec{Y}\rangle\}^2 - 4\left\|\vec{X}\right\|^2\left\|\vec{Y}\right\|^2 \le 0 \tag{2-25}$$

or

$$\left|\text{Re}\{\langle\vec{X},\vec{Y}\rangle\}\right| \le \left\|\vec{X}\right\|\left\|\vec{Y}\right\|. \tag{2-26}$$

Now, define the complex-valued constant

$$c = \overline{\langle\vec{X},\vec{Y}\rangle}\Big/\left|\langle\vec{X},\vec{Y}\rangle\right|, \tag{2-27}$$

where $\overline{\langle \vec{X}, \vec{Y} \rangle}$ denotes the complex conjugate of the inner product $\langle \vec{X}, \vec{Y} \rangle$. Now, from (2-26), observe that $\left|\text{Re}\{\langle c\vec{X}, \vec{Y} \rangle\}\right| \le \|c\vec{X}\|\,\|\vec{Y}\|$ or

$$\left|\text{Re}\{c\langle \vec{X}, \vec{Y} \rangle\}\right| \le |c|\,\|\vec{X}\|\,\|\vec{Y}\|. \tag{2-28}$$

But note that

$$|c| = \left|\overline{\langle \vec{X}, \vec{Y} \rangle}\right| \Big/ \left|\langle \vec{X}, \vec{Y} \rangle\right| = 1$$

$$c\langle \vec{X}, \vec{Y} \rangle = \left|\langle \vec{X}, \vec{Y} \rangle\right|^2 \Big/ \left|\langle \vec{X}, \vec{Y} \rangle\right| = \left|\langle \vec{X}, \vec{Y} \rangle\right|$$

so that (2-28) yields $\left|\langle \vec{X}, \vec{Y} \rangle\right| \le \|\vec{X}\|\,\|\vec{Y}\|$ as claimed.♥

The Cauchy-Schwarz inequality goes by various names, in the literature. Some authors call it the Cauchy Inequality while others call it the Schwarz Inequality (the Russian literature assigns the name(s) of one or more Russian mathematicians to the inequality).

The Cauchy-Schwarz inequality is used in many applications in engineering and the applies sciences. For example, the set of all real-valued random variable with finite second moments is a vector space (it is an example of a *Hilbert Space*). For this vector space or random variables, we define $\langle \vec{X}, \vec{Y} \rangle = E[\vec{X}\vec{Y}]$ and $\|\vec{X}\| = \sqrt{E[\vec{X}^2]}$ so that the Cauchy-Schwarz inequality becomes

$$\left|E[\vec{X}, \vec{Y}]\right|^2 \le E[\vec{X}^2]\,E[\vec{Y}^2], \tag{2-29}$$

a standard inequality given in most probability books.

Angle Between Vectors

The notion of dot product in $\mathcal{R}^2$ and $\mathcal{R}^3$ is discussed in most books on elementary statics and/or physics. In these books, the dot product of vectors $\vec{X}$ and $\vec{Y}$ is defined as

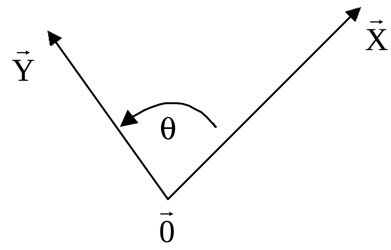

Figure 2-4: The angle between two vectors.

$$\left\langle \vec{X}, \vec{Y} \right\rangle \equiv \|\vec{X}\|_2 \|\vec{Y}\|_2 \cos\theta, \tag{2-30}$$

where θ is the angle between the vectors. We generalize this notion to $\mathcal{R}^n$, and define the *angle* between $\vec{X} \in \mathcal{R}^n$ and $\vec{Y} \in \mathcal{R}^n$ as the quantity θ which satisfies

$$\cos\theta \equiv \frac{\left\langle \vec{X}, \vec{Y} \right\rangle}{\|\vec{X}\|_2 \|\vec{Y}\|_2}. \tag{2-31}$$

By the Cauchy-Schwarz inequality, we know that the right-hand-side of this expression is equal to, or less than, unity, so cosθ exists. Furthermore, the angle θ is not changed if $\vec{X}$ and $\vec{Y}$ are replaced by $\alpha\vec{X}$ and $\beta\vec{Y}$, where α and β are any non-zero real numbers. Usually, angle θ is not defined for vector spaces that employ the field of complex numbers (since the right-hand side of (2-31) would be complex-valued, in general).

MatLab's Norm Function

MatLab supports the p-norm. The syntax for the p-norm is `norm(X,p)`. In MatLab, `norm(X,inf)` yields the infinity norm. The default is the 2-norm; the MatLab statement `norm(X)` is interpreted as `norm(X,2)`.

Gram-Schmidt Procedure

Let $\vec{X}_1$, $\vec{X}_2$, ... , $\vec{X}_n$, be n linearly independent vectors in an n-dimensional space. These vectors can form a basis for the space. They can be used to form n mutually orthogonal vectors

$\vec{Y}_1$, $\vec{Y}_2$, ... , $\vec{Y}_n$ by using the *Gram-Schmidt Procedure*. To form $\vec{Y}_k$, you start with $\vec{X}_k$, and subtract out the component in each of the directions $\vec{X}_1$, $\vec{X}_2$, ... , $\vec{X}_{k-1}$. The formulas for this are

$$
\begin{aligned}
\vec{Y}_1 &= \vec{X}_1 \\
\vec{Y}_2 &= \vec{X}_2 - \left\langle \vec{X}_2, \vec{Y}_1 \big/ \|Y_1\|_2 \right\rangle \frac{\vec{Y}_1}{\|Y_1\|_2} \\
\vec{Y}_3 &= \vec{X}_3 - \left\langle \vec{X}_3, \vec{Y}_1 \big/ \|Y_1\|_2 \right\rangle \frac{\vec{Y}_1}{\|Y_1\|_2} - \left\langle \vec{X}_3, \vec{Y}_2 \big/ \|Y_2\|_2 \right\rangle \frac{\vec{Y}_2}{\|Y_2\|_2} \\
&\ \vdots \qquad\qquad \vdots \\
\vec{Y}_n &= \vec{X}_n - \left\langle \vec{X}_n, \vec{Y}_1 \big/ \|Y_1\|_2 \right\rangle \frac{\vec{Y}_1}{\|Y_1\|_2} - \left\langle \vec{X}_n, \vec{Y}_2 \big/ \|Y_2\|_2 \right\rangle \frac{\vec{Y}_2}{\|Y_2\|_2} - \cdots - \left\langle \vec{X}_n, \vec{Y}_{n-1} \big/ \|Y_{n-1}\|_2 \right\rangle \frac{\vec{Y}_{n-1}}{\|Y_{n-1}\|_2}
\end{aligned}
\tag{2-32}
$$

This procedure can be modified easily to produce $\vec{Y}_1$, $\vec{Y}_2$, ... , $\vec{Y}_n$ that are orthonormal (orthogonal *and* unit length). The formulas are

$$
\begin{aligned}
\vec{Y}_1 &= \vec{X}_1 \big/ \|\vec{X}_1\|_2 \\
\vec{Y}_2 &= \frac{\vec{X}_2 - \langle \vec{X}_2, \vec{Y}_1 \rangle \vec{Y}_1}{\left\| \vec{X}_2 - \langle \vec{X}_2, \vec{Y}_1 \rangle \vec{Y}_1 \right\|_2} \\
\vec{Y}_3 &= \frac{\vec{X}_3 - \langle \vec{X}_3, \vec{Y}_1 \rangle \vec{Y}_1 - \langle \vec{X}_3, \vec{Y}_2 \rangle \vec{Y}_2}{\left\| \vec{X}_3 - \langle \vec{X}_3, \vec{Y}_1 \rangle \vec{Y}_1 - \langle \vec{X}_3, \vec{Y}_2 \rangle \vec{Y}_2 \right\|_2} \\
&\ \vdots \qquad\qquad \vdots \\
\vec{Y}_n &= \frac{\vec{X}_n - \langle \vec{X}_n, \vec{Y}_1 \rangle \vec{Y}_1 - \langle \vec{X}_n, \vec{Y}_2 \rangle \vec{Y}_2 - \cdots - \langle \vec{X}_n, \vec{Y}_{n-1} \rangle \vec{Y}_{n-1}}{\left\| \vec{X}_n - \langle \vec{X}_n, \vec{Y}_1 \rangle \vec{Y}_1 - \langle \vec{X}_n, \vec{Y}_2 \rangle \vec{Y}_2 - \cdots - \langle \vec{X}_n, \vec{Y}_{n-1} \rangle \vec{Y}_{n-1} \right\|_2}
\end{aligned}
\tag{2-33}
$$

In an n-dimensional space, let $\vec{X}_1$, $\vec{X}_2$, ... , $\vec{X}_r$, where r < n, be a mutually orthonormal set of r vectors. Then, this set can be extended to an orthonormal basis of the vector space.

1) Extend the set to a basis. Find $\vec{X}_{r+1}$, ... , $\vec{X}_n$ so that the n vectors $\vec{X}_1$, $\vec{X}_2$, ... , $\vec{X}_r$, $\vec{X}_{r+1}$, ... , $\vec{X}_n$ are independent.

2) Apply Gram-Schmidt, in the form of (2-33), to obtain an orthonormal basis this set of n vectors $\vec{Y}_1$, $\vec{Y}_2$, ... , $\vec{Y}_n$.

Example

Consider the following basis of $\mathcal{R}^3$

$$\vec{X}_1 = \begin{bmatrix}1\\1\\1\end{bmatrix}, \quad \vec{X}_2 = \begin{bmatrix}0\\1\\1\end{bmatrix}, \quad \vec{X}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}.$$

$$\vec{Y}_1 = \vec{X}_1 \Big/ \|\vec{X}_1\|_2 = \begin{bmatrix}1/\sqrt{3}\\1/\sqrt{3}\\1/\sqrt{3}\end{bmatrix}$$ is the first normalized vector.

Compute the intermediate vector $\vec{U}_2 \equiv \vec{X}_2 - \langle \vec{X}_2, \vec{Y}_1 \rangle \vec{Y}_1 = \begin{bmatrix}0\\1\\1\end{bmatrix} - \frac{2}{\sqrt{3}}\begin{bmatrix}1/\sqrt{3}\\1/\sqrt{3}\\1/\sqrt{3}\end{bmatrix} = \begin{bmatrix}-2/3\\1/3\\1/3\end{bmatrix}$ so that

$$\vec{Y}_2 = \vec{U}_2 \Big/ \|\vec{U}_2\|_2 = \begin{bmatrix}-2/\sqrt{6}\\1/\sqrt{6}\\1/\sqrt{6}\end{bmatrix}$$ is the second normalized vector. Compute the intermediate vector

$$\vec{U}_3 \equiv \vec{X}_3 - \langle \vec{X}_3, \vec{Y}_1 \rangle \vec{Y}_1 - \langle \vec{X}_3, \vec{Y}_2 \rangle \vec{Y}_2 = \begin{bmatrix}0\\0\\1\end{bmatrix} - \frac{1}{\sqrt{3}}\begin{bmatrix}1/\sqrt{3}\\1/\sqrt{3}\\1/\sqrt{3}\end{bmatrix} - \frac{1}{\sqrt{6}}\begin{bmatrix}-2/\sqrt{6}\\1/\sqrt{6}\\1/\sqrt{6}\end{bmatrix} = \begin{bmatrix}0\\-1/2\\1/2\end{bmatrix}$$ so that

$$\vec{Y}_3 = \frac{\vec{U}_3}{\|\vec{U}_3\|} = \begin{bmatrix}0\\-1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix}$$ is the third normalized vector.♥

Linear Transformations

Let n-dimensional **U** and m-dimensional **V** be vector spaces over the same field $\mathcal{F}$. A *linear transformation* **T** of **U** into **V** is a *single valued* mapping of **U** into **V** which associates with each $\vec{X} \in$ **U** a *unique* $\vec{Y} = \mathbf{T}(\vec{X}) \in$ **V** such that for all $\vec{X}_1$ and $\vec{X}_2$ in **U** and all scalars α_1 and α_2 in $\mathcal{F}$ we have

$$\mathbf{T}(\alpha_1\vec{X}_1 + \alpha_2\vec{X}_2) = \alpha_1\mathbf{T}(\vec{X}_1) + \alpha_2\mathbf{T}(\vec{X}_2). \tag{2-34}$$

For any $\vec{X} \in$ **U**, we call $\vec{Y} = \mathbf{T}(\vec{X})$ the *image of* $\vec{X}$ *under* T. For any $\vec{Y} \in$ **V** we call $\{\vec{X} \in \mathbf{U} : \vec{Y} = \mathbf{T}(\vec{X})\}$ the *inverse image* of $\vec{Y}$.

Generally, the inverse image of $\vec{Y}$ can be multidimensional (*i.e.*, Figure 2-5b is OK); there *may be* more than one $\vec{X} \in$ **U** that maps to a given $\vec{Y} \in$ **V**. However, transformation **T : U** $\rightarrow$ **V** is said to be ***one-to-one*** if for *all* $\vec{X}_1 \neq \vec{X}_2$ we have $\mathbf{T}(\vec{X}_1) \neq \mathbf{T}(\vec{X}_2)$ (*i.e.*, if **T** is ***one-to-one***, then Figure 2-5b never happens).

Let A be a subset of vector space **U**. **T**(A) denotes the set of all images of elements of A.

$$\mathbf{T}(A) = \{\vec{Y} \in \mathbf{V} : \vec{Y} = \mathbf{T}(\vec{X}),\ \vec{X} \in A\} \tag{2-35}$$

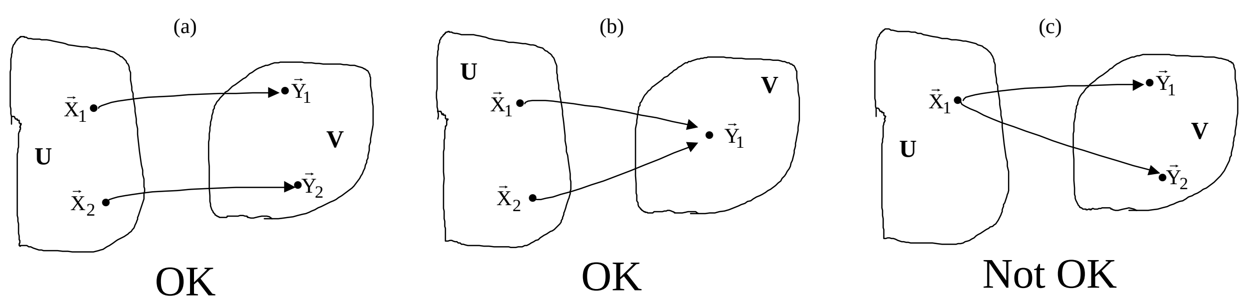

Figure 2-5: Linear Transformation of U into V. Mappings a) and b) are, and mapping c) is not, permissible.

$\mathbf{T}(A)$ is called the *image of A*.

If $\mathbf{T}(\mathbf{U}) = \mathbf{V}$, we say that **T** is a mapping from **U** *onto* **V**. **T** is a mapping from **U** *onto* **V** if for each $\vec{Y} \in \mathbf{V}$ there exists an $\vec{X} \in \mathbf{U}$ such that $\vec{Y} = \mathbf{T}(\vec{X})$.

Suppose $\mathbf{T} : \mathbf{U} \rightarrow \mathbf{V}$ is both *one-to-one* and *onto.* Then, **T** has an inverse, denoted as $\mathbf{T}^{-1}$, and $\mathbf{T}^{-1}$ is a linear transformation mapping **V** onto **U** ($\mathbf{T}^{-1} : \mathbf{V} \rightarrow \mathbf{U}$). If $\mathbf{T} : \mathbf{U} \rightarrow \mathbf{V}$ is not both *one-to-one* and *onto*, then an inverse does not exist.

Regardless of whether or not **T** is one-to-one or onto, $\mathbf{T}(\mathbf{U})$ is a subspace called the *range* of **T**. We write $R(\mathbf{T})$ to denote the range of **T** (note that $R(\mathbf{T}) = \mathbf{T}(\mathbf{U})$; however, we will use $R(\mathbf{T})$ to denote the range of **T**). Symbolically, the range of **T** can be expressed as

$$R(\mathbf{T}) = \left[\vec{Y} = \mathbf{T}(\vec{X}) : \vec{X} \in \mathbf{U}\right] \tag{2-36}$$

The dimension of $R(\mathbf{T})$ is called the *rank* of **T**. Let $n = \dim(\mathbf{U})$ and $m = \dim(\mathbf{V})$. Then

$$\text{Rank of } \mathbf{T} \le \min(m, n) \tag{2-37}$$

The *kernel* of $\mathbf{T} : \mathbf{U} \rightarrow \mathbf{V}$, denoted as $K(\mathbf{T})$, contains all the vectors in **U** that map to $\vec{0}$; that is,

$$K(\mathbf{T}) = \{\vec{X} \in \mathbf{U} : \mathbf{T}(\vec{X}) = \vec{0}\}. \tag{2-38}$$

The *kernel* is a subspace of **U,** and the dimension of $K(\mathbf{T})$ is called the *nullity* of **T**.

Each $\vec{X} \in \mathbf{U}$ maps to a non-zero vector in the $R(\mathbf{T})$, or it is in $K(\mathbf{T})$. Hence, it is easy to argue that

$$\text{Rank}(\mathbf{T}) + \text{Nullity}(\mathbf{T}) = n = \dim(\mathbf{U}) \tag{2-39}$$

Also argued easily are the facts that $\mathbf{T} : \mathbf{U} \rightarrow \mathbf{V}$ is one-to-one if and only if Nullity($\mathbf{T}$) = 0, and $\mathbf{T} : \mathbf{U} \rightarrow \mathbf{V}$ is onto if and only if Rank($\mathbf{T}$) = m = dim($\mathbf{V}$).

Special Cases

1) **dim(U) = n < dim(V) = m**. Then Rank ($\mathbf{T}$) = n - Nullity($\mathbf{T}$) $\leq$ n < m so **T** cannot be onto. If **T** transforms a space to a higher dimensional space then it cannot be onto.

2) **dim(U) = n > dim(V) = m**. Then Nullity ($\mathbf{T}$) = n - Rank($\mathbf{T}$) $\geq$ n - m > 0 so **T** cannot be one-to-one. If **T** transforms a space to a lower dimensional space then it cannot be one-to-one.

Existence and Uniqueness of Linear Transformations

Let $\vec{\alpha}_1, \vec{\alpha}_2, \ldots, \vec{\alpha}_n$ be a basis of n dimensional **U**. Let $\vec{Y}_1, \vec{Y}_2, \ldots, \vec{Y}_n$ be *any* vectors in **V** (not necessarily independent). Then there exists a *unique* linear transformation $\mathbf{T} : \mathbf{U} \rightarrow \mathbf{V}$ such that $\mathbf{T}(\vec{\alpha}_k) = \vec{Y}_k$, $1 \leq k \leq n$. Here, the independence of the vectors $\vec{\alpha}_1, \vec{\alpha}_2, \ldots, \vec{\alpha}_n$ is crucial in establishing the *existence* of **T** with the desired properties. A linear dependence among the $\vec{\alpha}_k$'s would impose a linear dependence on the $\mathbf{T}(\vec{\alpha}_k)$'s so that the $\vec{Y}_k$'s could not be specified arbitrarily.

Let $\vec{\alpha}_1, \vec{\alpha}_2, \ldots, \vec{\alpha}_r$ be linearly independent vectors in **U**, and assume that r < n = dim(**U**). Let $\vec{Y}_1, \vec{Y}_2, \ldots, \vec{Y}_r$ be *any* vectors in **V** (not necessarily independent). Then there exists a linear transformation $\mathbf{T} : \mathbf{U} \rightarrow \mathbf{V}$ such that $\mathbf{T}(\vec{\alpha}_k) = \vec{Y}_k$, $1 \leq k \leq r$. However, the transformation may not be unique.